Detecting Multiple Change-Points in the Mean of Gaussian Process by Model Selection

Emilie Lebarbier

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Detecting Multiple Change-Points in the Mean of Gaussian Process by Model Selection

Emilie Lebarbier

N° 4740

Février 2003

THÈME 4

ISSN 0249-6399 ISRN INRIA/RR--4740--FR+ENG

INRIA
RHÔNE-ALPES

Detecting Multiple Change-Points in the Mean of Gaussian Process by Model Selection

Emilie Lebarbier *

Thème 4 — Simulation et optimisation
de systèmes complexes

Projet IS2

Rapport de recherche n° 4740 — Février 2003 — 24 pages

Abstract: This paper deals with the problem of detecting the change-points in mean of a signal corrupted by an additive Gaussian noise. The number of changes and their positions are unknown. From a nonasymptotic point of view, we propose to estimate them with a method based on a penalized least-squares criterion. According to the results of Birgé and Massart, we choose the penalty function such that the resulting estimator minimizes the quadratic risk. This penalty depends on unknown constants and we propose a calibration leading to an automatic method. The performances of the method are assessed through simulation experiments. An application to real data is shown.

Key-words: Detection of change-points; Penalized contrast; Model selection

* Emilie.Lebarbier@inrialpes.fr. This research has been realized when E.Lebarbier was with Laboratoire de mathématiques - Université Paris XI.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 Montbonnot-St-Martin (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Détection de ruptures dans la moyenne d'un processus gaussien par une méthode de sélection de modèle

Résumé : Le papier traite du problème de détection de ruptures dans la moyenne d'un signal gaussien. Le nombre de ruptures et leurs localisations sont supposés inconnus. D'un point de vue non-asymptotique, nous proposons de les estimer à l'aide d'une méthode basée sur un critère des moindres carrés pénalisés. En appliquant les résultats de Birgé and Massart, nous choisissons une fonction de pénalité telle que l'estimateur pénalisé correspondant réalise le risque quadratic minimal. Cette pénalité dépend de constantes inconnues et nous proposons de les calibrer afin d'obtenir une automatique méthode. Une étude de simulations est menée pour évaluer la performance de la méthode, et une application sur des données réelles est réalisé.

Mots-clés : Détection de ruptures ; Contraste pénalisé ; Sélection de modèles

1 Introduction

The following change-points in the mean model is considered:

$$y_t = s(x_t) + \varepsilon_t \quad t = 1, \ldots, n \tag{1.1}$$

where $x_t = \frac{t}{n}$ and the errors (ε_t) are supposed to be zero-mean, identically distributed unobservable Gaussian independant random variables of common variance σ^2. The function s to be recovered is assumed to be piecewise constant. Thus, there exists some instants $\tau_0 = 0 < \tau_1 < \ldots < \tau_K = 1$ such that the function s is constant between two successive change-points instants. In other words, there exists a sequence $(s_1, \ldots, s_K)$ such that, for any $k \geq 1$,

$$s = \sum_{k=1}^{K} s_k \mathbb{1}_{I_k} \quad \text{with } I_k =]\tau_{k-1}, \tau_k], \tag{1.2}$$

This model means that $K-1$ changes affect the mean of (y_t) at some unknown instants $(t_k, 1 \leqslant k \leqslant K-1)$ with $t_k = [n\tau_k]$ where the number of change-points $(K-1)$ is supposed to be unknown. The problem is to detect and locate the change-points instants of y and to estimate the jumps of mean.

The change detection problem has been studied for more than fourteen years in many frameworks. One can refer to the books of Basseville and Nikiforov [2], Brodsky and Darkhovsky [6], Carlstein, Müller and Siegmund [9] for a complete bibliography.
In the case of multiple changes, the problem is more intricate and few approaches are dedicated to this problem. Among them Bayesian techniques [15], and penalized criteria [17], [16], [8] can be distinguished. This kind of criterion is now classical (the first examples are the Mallows'C_p within the framework of the regression [20], the AIC [1] and the BIC [22] criteria within the framework of maximum likelihood estimation) and has been studied by many authors in the particular problem of detecting change-points in the mean, see for example Yao [23], Miao *et al.* [21] and more recently Lavielle et Moulines [17]. Their purpose is to estimate consistently all the change-points. In theory that requires the function s to be in the collection of models considered for fitting and in practice the penalty must be chosen according to practical considerations. Contrary to this point of view, we consider in this paper a nonasymptotic approach. The problem is seen as a particular problem of fixed design Gaussian regression where the regression function s to be estimated is piecewise constant. We consider the nonparametric model selection point of view developed by Birgé and Massart [3]. Their aim is to choose s minimizing a quadratic risk, by using as few prior information as possible, rather than to determine the true function s. This is a situation where it may be preferable to ignore some change-points corresponding to small jumps of mean. While applying their results, we give in Proposition 2.1 the penalty function involved in this approach and an upper bound of the quadratic risk for the corresponding penalized estimator.
The penalty function depends on two constants whose optimal values are not accessible theoretically, and on the noise variance σ^2 which is unknown in practice. We propose a two-step procedure to estimate the penalty. First, the noise variance is supposed to be known and the values of the two constants are calibrated by adapting the simulation procedure proposed by Birgé and Rozenholc [5] in the density estimation framework by histograms in model selection approach studied by Castellan [7]. Then, the noise variance is seen as a penalty constant and it is estimated with a method based on heuristic and theoretical ideas proposed by Birgé and Massart [4]. This method is detailled in this paper. According to some difficulties occured

on simulated data with this method, a specific calibration is proposed.
In numerical experiments, our calibrated method is compared to the C_p Mallows [20] and the BIC criteria, see Schwarz [22], for which asymptotic properties have been etablished. Then two others simulations studies are performed to assess the method when the model assumptions are not valid. In the first one, the function s is not a piecewise constant function and in the second one the noise is not Gaussian.

The paper is organized as follows. In Section 2 the estimation procedure is described and the form of the penalty function and an upper bound of the quadratic risk for the corresponding penalized estimator are given. When the variance of the noise is known, the two penalty constants are calibrated in a optimal way in Section 3. Then the heuristic method to estimate the noise variance is presented and a calibration of this method is proposed in Section 4. Section 5 shows different simulations studies to assess the performance of our calibrated method. Finally the method is applied to detect change-points in the monthly number of tests HIV in France on several years. A short discussion ends the paper.

2 Estimation procedure

In this section, the estimation procedure is presented. The principle is the following one: A collection of least-squares estimators is designed and the best according to a penalized least-squares criterion is chosen. The form of the penalty function and a nonasymptotic risk bound for the corresponding penalized estimator are given.

2.1 Collection of estimators and ideal one

Let $\mathcal{M}_n$ be the set of all the partitions on the grid $\{x_1, x_2, \ldots, x_n\}$. There is the need to visit all partitions as possible since no prior information on the location of the change-points instants is available. We consider the collection $\{\mathcal{S}_m, m \in \mathcal{M}_n\}$ where for a given partition m of dimension D_m, $m = \{I_k\}_{k=1,\ldots,D_m}$, $\mathcal{S}_m$ is the linear subspace of the functions which are piecewise constant on the partition m,

$$\mathcal{S}_m = \left\{ u = \sum_{k=1}^{D_m} u_k \mathbb{1}_{I_k}, (u_k)_{k=1,\ldots,D_m} \in \mathbb{R}^{D_m} \right\}.$$

Denoting $\mathcal{S} = \bigcup_{m \in \mathcal{M}_n} \mathcal{S}_m$, we consider the least-squares criterion, defined for any $u \in \mathcal{S}$,

$$\gamma_n(u) \quad = \quad \frac{1}{n} \sum_{t=1}^{n} [y_t - u(x_t)]^2 = \|y - u\|_n^2,$$

where $\|.\|_n$ denotes the normalized Euclidean norm on $\mathbb{R}^n$. Associated to the collection $\{\mathcal{S}_m, m \in \mathcal{M}_n\}$, we construct the collection of least-squares estimators $\{\hat{s}_m, m \in \mathcal{M}_n\}$ where for $m \in \mathcal{M}_n$:

$$\hat{s}_m \quad = \quad \underset{u \in \mathcal{S}_m}{\operatorname{argmin}} \, \gamma_n(u) = \sum_{k=1}^{D_m} \bar{y}_k \mathbb{1}_{I_k},$$

where $\bar{y}_k$ is the empirical mean of y on the interval I_k, *i.e.* $\bar{y}_k = \frac{1}{n_k} \sum_{t=t_{k-1}+1}^{t_k} y_t$, with n_k is the number of x_t belonging to the interval I_k and let recall that $t_k = [n\tau_k]$.

The estimation problem of s reduces to choose the best partition, say $\hat{m}$ and take $\hat{s}_{\hat{m}}$. Considering the loss function, which is associated to the empirical contrast function γ_n for any $u \in \mathcal{S}$ by the relation

$$l(s,u) \quad = \quad \mathbb{E}_s\left[\gamma_n(u) - \gamma_n(s)\right] = \|s-u\|_n^2, \tag{2.3}$$

and taking the loss mean of the least-squares estimator $\hat{s}_m$, leads to its quadratic risk which is decomposed as follows:

$$\mathbb{E}_s\left[\|s-\hat{s}_m\|_n^2\right] = \|s-\bar{s}_m\|_n^2 + \frac{D_m}{n}\sigma^2, \tag{2.4}$$

where $\bar{s}_m$ is the orthogonal projection of s on $\mathcal{S}_m$:

$$\bar{s}_m = \underset{u\in\mathcal{S}_m}{\operatorname{argmin}}\ l(s,u) = \sum_{k=1}^{D_m}\left(\frac{1}{n_k}\sum_{t=t_{k-1}+1}^{t_k} s(x_t)\right)\mathbb{1}_{I_k}.$$

In this decomposition (2.4), the term $\|s-\bar{s}_m\|_n^2$ is a bias term measuring the quality of approximation of s by $\mathcal{S}_m$, whereas the term $\frac{D_m}{n}\sigma^2$ is a variance term representing the estimation error in $\mathcal{S}_m$.

The ideal partition, say $\bar{m}(s)$ is minimizing the risk (2.4) over $\mathcal{M}_n$,

$$O_n(s,\mathcal{S}) = \inf_{m\in\mathcal{M}_n}\mathbb{E}_s\left[\|s-\hat{s}_m\|_n^2\right]. \tag{2.5}$$

Unfortunately, $\bar{m}(s)$ is unknown since it depends on the unknown function s. The purpose of the proposed estimation procedure is to provide a data-driven criterion selecting an estimator $\tilde{s}$ having a risk as close as possible to the risk of $\hat{s}_{\bar{m}(s)}$, namely such that

$$\mathbb{E}_s[l(s,\tilde{s})] \leqslant \quad C \quad O_n(s,\mathcal{S}), \tag{2.6}$$

for a nonnegative constant C.

2.2 The proposed estimator

Given some penalty function $pen_n(m) \ : \mathcal{M}_n \to \mathbb{R}^+$, the penalized least-sqares estimator is defined by:

$$\tilde{s} = \hat{s}_{\hat{m}}, \tag{2.7}$$

where $\hat{m}$ minimizes the penalized least-squares criterion over $\mathcal{M}_n$ defined by:

$$crit_n(m) = \gamma_n(\hat{s}_m) + pen_n(m). \tag{2.8}$$

The problem is to choose a convenient penalty function which selects $\tilde{s}$ leading to an inequality of the type (2.6). We will see that the choice of this penalty function depends on the richness of the collection of partitions $\mathcal{M}_n$.

The following proposition gives the form of the penalty function and a risk bound of the associated penalized estimator.

Proposition 2.1. *There exists two positive constants c_1 and c_2 such that if the penalty is defined for all partition $m \in \mathcal{M}_n$ by*

$$pen_n(m) = \frac{D_m}{n}\sigma^2\left(c_1\log\left(\frac{n}{D_m}\right) + c_2\right), \tag{2.9}$$

then there exists some constants $C(c_1, c_2)$ et $C'(c_1, c_2)$ such that the risk of the penalized estimator $\tilde{s}$, defined in (2.7), satisfies

$$\begin{aligned}\mathbb{E}_s[l(s,\tilde{s})] &\leqslant C(c_1,c_2) \inf_{m\in\mathcal{M}_n}\left[\|s-\bar{s}_m\|_n^2 + pen_n(m)\right] \\ &\quad + C'(c_1,c_2)\frac{\sigma^2}{n}. \end{aligned} \tag{2.10}$$

Proof. The problem can be embedded into the Gaussian process framework and we use a result of Birgé and Massart [3]. First we recall the theorem giving the general form of the penalty function and a risk bound of the associated penalized estimator.

Theorem 2.2. *(Birgé, Massart) [3]. Let $\{L_m\}_{m\in\mathcal{M}_n}$ be a family of weights, i.e. nonnegative real numbers, satisfying the condition*

$$\Sigma = \sum_{\{m\in\mathcal{M}|D_m>0\}} e^{-L_m D_m} < +\infty$$

Let us then consider a penalty function such that

$$pen(m) \geq K\sigma^2\frac{D_m}{n}(1+\sqrt{2L_m})^2 \quad \textit{for all } m\in\mathcal{M}_n \textit{ and some } K>1, \tag{2.11}$$

the corresponding penalized estimator $\tilde{s}$ exists almost already and is unique. Moreover it satisfies

$$\mathbb{E}_s[l(s,\tilde{s})] \leqslant C(K) \inf_{m\in\mathcal{M}_n}[l(s,\bar{s}_m) + pen(m)] + C'(K)\frac{\Sigma}{n}\sigma^2.$$

We need to choose a convenient family of weights $\{L_m\}_{m\in\mathcal{M}_n}$. Taking L_m as a function of the dimension, $L_m = L(D_m) = L_{D_m}$ leads to

$$\begin{aligned}\Sigma &= \sum_{m\in\mathcal{M}} e^{-L_m D_m} = \sum_{D=1}^{n} e^{-DL_D} Card\{m\in\mathcal{M}_n, D_m = D\} \\ &\leqslant \sum_{D=1}^{n} e^{-DL_D}\binom{n}{D} \\ &\leqslant \sum_{D=1}^{n} e^{-DL_D}\left(\frac{en}{D}\right)^D \\ &\leqslant \sum_{D=1}^{n} e^{-D\left(L_D - 1 - \log\left(\frac{n}{D}\right)\right)}. \end{aligned}$$

Consequently, if we take $L_D = 2 + \log\left(\frac{n}{D}\right)$ then $\Sigma < 1$ and this leads to a penalty function of the form (2.9). Then we get Proposition 2.1 deduced from Theorem 2.2. ■

Remark 1. The penalty depends on the partition m only via its dimension D_m. The factor $\log\left(\frac{n}{D_m}\right)$ appearing in this penalty results from the complexity of the considered collection of partitions $\mathcal{M}_n$, *i.e.* the number of partitions having the same dimension in this collection (here $\binom{n-1}{D-1}$ for a fixed dimension D). This

factor can be regarded as surprising: for instance, it does not appear in the penalty of the C_p Mallows's criterion [20]. Birgé and Massart [3] show that this term is necessary in the risk given by (2.10) from a minimax point of view.

Remark 2. Note that the considered collection of partitions, $\mathcal{M}_n$, depends on the size of the sample n contrary to the Cp Mallows [20] or AIC heuristics [1].

Remark 3. According to the inequality (2.10), the risk of the penalized estimator is bounded by

$$\mathbb{E}_s[l(s,\tilde{s})] \leqslant C \log(n)\, O_n(s,\mathcal{S}). \tag{2.12}$$

That means that, in terms of risk, $\tilde{s}$ is as good as the best of the $\hat{s}_m$ with a factor $\log n$.

In the theoretical approach, the penalty function is depending on a known variance σ^2. However in practical implementation the variance must be estimated. Moreover the penalty function depends on two constants c_1 and c_2 which optimal values are unknown. To deal with both problems, we proceed in two steps. First in Section 3, the variance of the noise σ^2 is supposed to be known and a calibration of the optimal values of c_1 and c_2 from a model selection point of view is proposed by a simulation procedure. Then in Section 4, with c_1 and c_2 fixed to the previous values, σ^2 is estimated by a heuristic method and a calibration of this method is proposed.

3 Choice of c_1 and c_2 in the penalty function

In this section, σ^2 is supposed to be known. Before describing the simulation procedure to approximate the optimal values of c_1 and c_2, we define an appropriate reference of quality of the penalized estimator since the classical one given by (2.5) appears to be unreliable in our framework.

3.1 What is the good reference of quality ?

Recall that the aim is to provide a penalty function such that the risk of the corresponding estimator, say $\tilde{s}(c_1,c_2)$, is as close as possible to the minimum one $O_n(s,\mathcal{S})$ (2.5). It is then natural to evaluate the performance of $\tilde{s}(c_1,c_2)$ by the measurement of the following ratio $\mathbb{E}_s[\|s-\tilde{s}(c_1,c_2)\|_n^2]/O_n(s,\mathcal{S})$. We are looking for the values of c_1 and c_2 that minimizes this risk ratio uniformly for all function s and sample size n. According to (2.12), this ratio is bounded by a quantity depending on n. Consequently the penalty can not be calibrated with respect to n.

We have chosen a new reference of quality to be the following one:

$$O_{(n,r)}(s,\mathcal{S}) = \inf_{D=1,\dots,n} \mathbb{E}_s[\|s-\hat{s}_D\|_n^2], \tag{3.13}$$

where $\hat{s}_D$ is the best least-squares estimator of dimension D:

$$\hat{s}_D = \operatorname*{argmin}_{\{u\in\mathcal{S}_D=\bigcup_{m\in\mathcal{M}_n,|m|=D}\mathcal{S}_m\}} \gamma_n(u) = \operatorname*{argmin}_{\{m\in\mathcal{M}_n,|m|=D\}} \gamma_n(\hat{s}_m).$$

since the penalty function depends on the partition only via its dimension. The least-squares estimators $\{\hat{s}_D, D = 1, \ldots, n\}$ are computed using a dynamic programming with a computationnal complexity of order of $\mathcal{O}(n^2)$ (we refer the reader for more details about this algorithm to the book of Kay [14]). The final estimator is $\tilde{s} = \hat{s}_{\hat{D}}$ where $\hat{D}$ minimizes the penalized criterion (2.8).
The performance of the penalized least-squares estimator is then measured with the risk ratio

$$F_n(s, c_1, c_2) = \frac{\mathbb{E}_s\left[\|s - \tilde{s}(c_1, c_2)\|_n^2\right]}{O_{(n,r)}(s, \mathcal{S})}. \tag{3.14}$$

This ratio seems to be more appropriate since it is expected to be bounded independently on s and n. This is true as soon as

$$\mathbb{E}_s[\|s - \hat{s}_D\|_n^2] = \inf_{\{m \in \mathcal{M}_n, |m| = D\}} \{\|s - s_m\|_n^2 + pen_n(m)\},$$

but we have no proof for that guess which has been confirmed by simulations in the next subsection.

3.2 Simulation procedure

We consider a collection of values of n denoted by $\mathcal{N} = \{20, 50, 100, 300, 500, 1000, 5000\}$ and a collection of 35 piecewise constant functions denoted by $\mathcal{L}$ and randomly simulated in the following way. We simulate

- the number of pieces $nseg = X + 1$ where X follows a Poisson distribution with parameter 5.
- each change-point instant τ_k, $k = 1, \ldots, nseg - 1$ follows an Uniform distribution on $]0, 1]$.
- each mean s_k, $k = 1, \ldots, nseg$ follows a standardized Gaussian distribution.

We fix $\sigma^2 = 1$ and the maximal dimension of the partition $D_{max} = 40$ since in practice it is not necessary to compute the partitions having a too large dimension.

We search for the values c_1 and c_2 performing as well as possible for all the functions on $\mathcal{L}$. It leads choosing c_1 and c_2 minimizing

$$F_n(c_1, c_2) = \sup_{s \in \mathcal{L}} F_n(s, c_1, c_2). \tag{3.15}$$

For any n, any s and several values of c_1 and c_2, $\mathbb{E}_s\left[\|s - \tilde{s}(c_1, c_2)\|_n^2\right]$ and $O_{(n,r)}(s, \mathcal{S})$ are evaluated by the empirical means obtained over 250 simulations. The supremum (3.15) is then calculate for n in $\mathcal{N}$. This leads to a large set of values $\{F_n(c_1, c_2), c_1, c_2 > 0, n \in \mathcal{N}\}$ summarized in Figure 1 by the functions $c_1 \to F_n(c_1, c_2)$ for three values of $c_2 = 0,\ 5,\ 8$.
Let $c_1^*(n, c_2)$ be the minimizer of $F_n(c_1, c_2)$. The idea is to declare as optimal value of the constant c_2, say c_2^*, the one which makes $c_1^*(n, c_2)$ stable with respect to n. The optimal value of c_1 will be $c_1^* = c_1^*(n, c_2^*)$ for any n.

Some comments are in order:

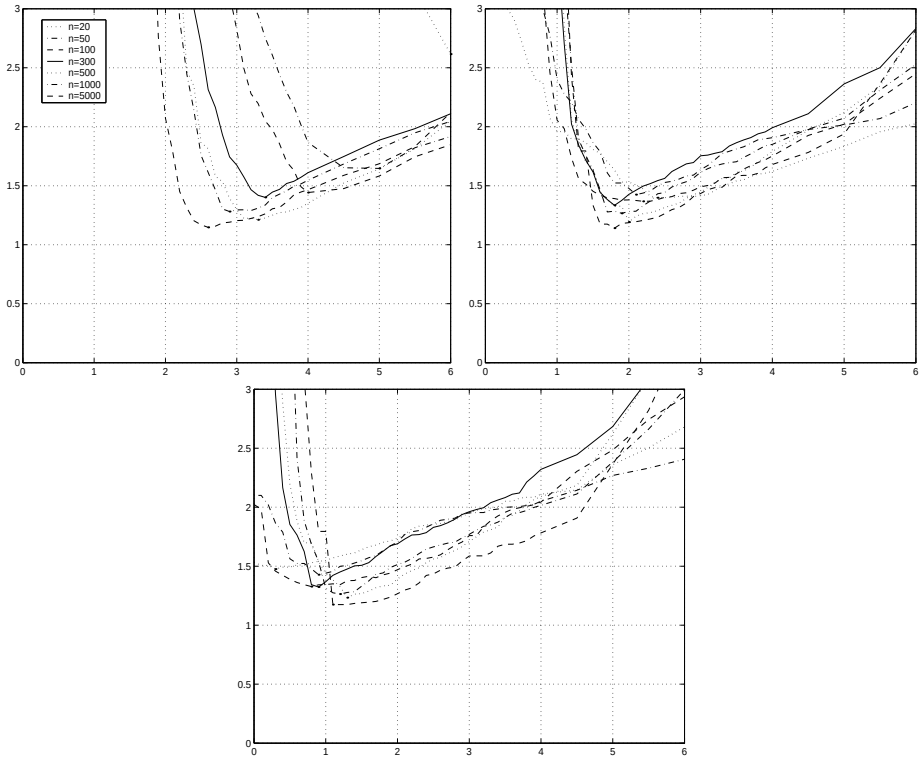

Figure 1: $c_1 \to F_n(c_1, 0)$ (in the left on the top), $c_1 \to F_n(c_1, 5)$ (in the right on the top) and $c_1 \to F_n(c_1, 8)$ for each n.

- $c_1^*(n, 0)$ decreases with n.
- $c_1^*(n, 8)$ tends to increase with n.
- for c_2 in $[4.8, 6]$, $c_1^*(n, c_2)$ seems to be stable around 2 for any n.

Table 1 summarizes numerical values of different risk ratios. For each n, it gives

- $F_n(c_1^*(n), c_2^*(n))$ where $(c_1^*(n), c_2^*(n))$ are the minimizers of $F_n(c_1, c_2)$ and $F_n(2, 5)$.
- the average of $F_n(c_1^*(n, s), c_2^*(n, s))$ over $s \in \mathcal{L}$ denoted by MF_n where $(c_1^*(n, s), c_2^*(n, s))$ are the minimizers of $F_n(s, c_1, c_2)$ and the average of $F_n(s, 2, 5)$ over s denoted by $MF_n^{(2,5)}$.

The risk ratios appear to be close. For example, the difference between $F_n(2, 5)$ and $F_n(c_1^*(n), c_2^*(n))$ for all n is lower than 0.132.

Thus, it seems to be reasonnable to choose

$$(c_1^*, c_2^*) = (2, 5).$$

Remark that the different risk ratios are smaller than 2 and tends to 1 with n. This result confirms the statement that the risk ratio $F_n(s, c_1, c_2)$ given by (3.14) is bounded independently of n and so that our choosen reference of quality given by (3.13) appears to be adequate in our framework.

	$F_n(c_1^*(n), c_2^*(n))$	$F_n(2,5)$	MF_n	$MF_n^{(2,5)}$
$n = 20$	1.397	1.53	1.0654	1.213
$n = 50$	1.424	1.477	1.0588	1.17
$n = 100$	1.3	1.379	1.0657	1.142
$n = 300$	1.31	1.426	1.0588	1.127
$n = 500$	1.185	1.193	1.022	1.082
$n = 1000$	1.26	1.28	1.026	1.077
$n = 5000$	1.132	1.186	1.024	1.048

Table 1: Estimation of different risk ratios for each n.

4 Estimation of the penalty constant

Since the constant c_1 and c_2 are now determined, only the noise variance σ^2 is unknown. A method based on theoretical and heuristics ideas proposed in Birgé and Massart [4] is considered. It is employed in Letue [19] for the adjustment of a regression function in the Cox model. In Subsection 4.1 the general method is presented and its calibration is proposed in Subsection 4.2.

4.1 The method

Consider the general penalty function

$$pen_{\beta,n}(D) = \beta f_n(D) \quad \text{for all } D \geq 1, \tag{4.16}$$

where f_n is a suitable penalizing function. The associated penalized criterion is

$$crit_{\beta,n}(D) = \gamma_n(\hat{s}_D) + pen_{\beta,n}(D). \tag{4.17}$$

The principle of the basic heuristic for choosing β is the following one: for large D, $\gamma_n(\hat{s}_D)$ is linear with respect to $f_n(D)$ and the estimated slope is $-\beta/2$. In Subsection 4.1.1, this heuristic is sketched and performed in Subsection 4.1.2. In this subsection, difficulties of this heuristic are described. In Subsection 4.1.3, an extension of it proposed by Birgé and Massart is presented.

4.1.1 The basic heuristic

Identifying (4.17) with the C_p Mallows criterion [20], which has the form

$$crit(D) \quad = \quad \gamma_n(\hat{s}_D) + 2\, \mathbb{E}_s[\|\hat{s}_D - \bar{s}_D\|_n^2], \tag{4.18}$$

leads to $\mathbb{E}_s[\|\hat{s}_D - \bar{s}_D\|_n^2] = \alpha f_n$ with $\beta = 2\alpha$. Moreover, we have

$$\begin{aligned} \mathbb{E}_s[\gamma_n(\hat{s}_D) - \gamma_n(s)] &= \mathbb{E}_s[\gamma_n(\bar{s}_D) - \gamma_n(s)] + \mathbb{E}_s[\gamma_n(\hat{s}_D) - \gamma_n(\bar{s}_D)] \\ &= \mathbb{E}_s[\gamma_n(\bar{s}_D) - \gamma_n(s)] - \alpha f_n(D). \end{aligned}$$

By simulation, $\mathbb{E}_s[\gamma_n(\bar{s}_D) - \gamma_n(s)]$ is of the order of $\mathbb{E}_s[\|\bar{s}_D - s\|_n^2]$ from a certain D. When the considered partition is high-dimensional, one can consider that the bias term, $\mathbb{E}_s[\|\bar{s}_D - s\|_n^2]$, is close to zero. If $\gamma_n(\hat{s}_D)$ is concentered around its expectation, then $\gamma_n(\hat{s}_D)$ is of the order of $-\alpha f_n(D)$ and the slope of $\gamma_n(\hat{s}_D)$ with

respect to $f_n(D)$, say $-\hat{\alpha}$, is an estimator of $-\alpha$. Finally the penalized estimator $\tilde{s}$ is $\hat{s}_{\hat{D}}$ where $\hat{D}$ minimizes the criterion $crit_{2\hat{\alpha},n}$ (4.17).

Remark 4. When the collection of partitions $\mathcal{M}_n$ has one partition per dimension, then $\mathbb{E}_s[\gamma_n(\bar{s}_D) - \gamma_n(s)]$ is equal to $\|\bar{s}_D - s\|_n^2$.

4.1.2 Application

The heuristic is applied on three particular realizations, chosen to show various situations, with

$$\alpha = \sigma^2, \quad \text{and} \quad f_n(D) = \frac{D}{n}\left(\log\left(\frac{n}{D}\right) + 2.5\right) \quad \text{for all } D \geq 1.$$

We will see that this heuristic is easy to apply as soon as $\gamma_n(\hat{s}_D)$ is linear with respect to $f_n(D)$.

Consider a function g plotted in Figure 2 (g). We set $\sigma^2 = 1$. Three realizations, noted $y(a)$, $y(b)$ and $y(c)$, are simulated from the three following cases and are plotted in Figure 2 :

- (a) : difficult detection and small n : $s = g$ and $n = 60$;
- (b) : difficult detection and large n : $s = g$ and $n = 300$;
- (c) : easy detection and small n : $s = 3g$ and $n = 60$.

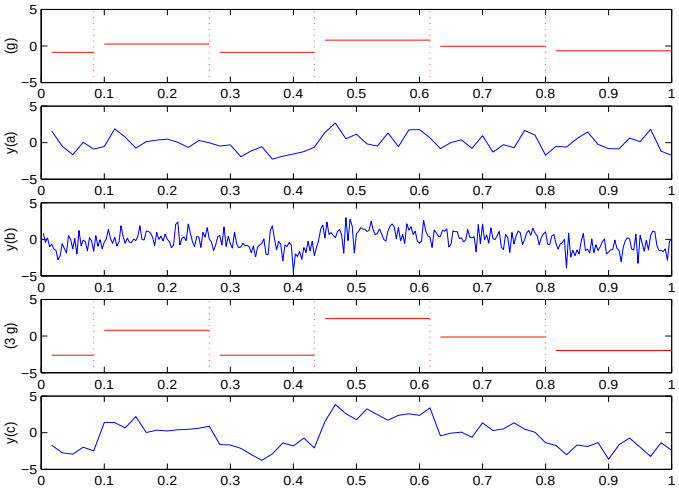

Figure 2: The three realizations and the functions from they are simulated.

The associated graph of $(f_n(D), \gamma_n(\hat{s}_D))$ are plotted in Figure 3. For large n, $\gamma_n(\hat{s}_D)$ is linear with respect to the function $f_n(D)$ in large dimensions (see Figure 3 (b)), and the choice of the points to estimate the regression coefficient is not sensitive. But this is not the case for small n (see Figure 3 (a)) for which $\gamma_n(\hat{s}_D)$ has rather a logarithmic behavior with respect to $f_n(D)$. Thus, the estimation of the regression coefficient is highly depending of the chosen interval. When the change-points are

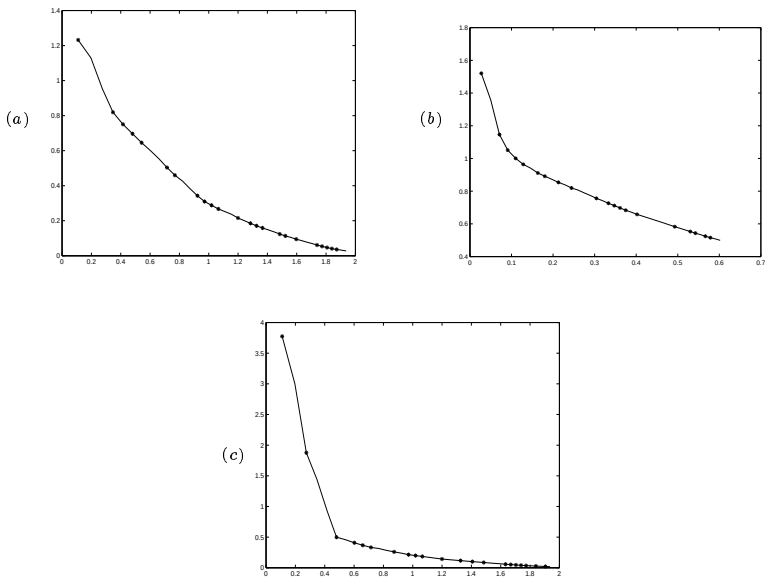

Figure 3: Graph of $(f_n(D), \gamma_n(\hat{s}_D))$ for $D = 1, \ldots, 40$ for the three realizations.

very sharp (see Figure 3 (c)), the graph shows a marked elbow for dimension 5. Remark that $L-curve$ methods (see for example [13], [12]...) can be applied in this particular configuration without any difficulty. A more complete simulation study done in [18] is showing the influence of the chosen dimensions to estimate the slope particulary for small n.

The question is then how to choose the dimension interval to get an honest estimate of the regression slope.

4.1.3 Selecting the dimension range values to apply the heuristic

An extension of the heuristic developed, by Birgé and Massart [4] is considered here. They show in [3] that a too small penalty value leads to select a high-dimensional partition ($K < 1$ for (2.11)). The penalty with $K > 1$ close to 1 is called minimal penalty. It is associated to the penalty $\alpha f_n(D)$ and leads to a reasonnable dimension. Their idea is to consider different values of α from 0 in $pen_{\alpha,n}$ (4.16) and select the associated dimension, say $\hat{D}(\alpha)$, minimizing $crit_{\alpha,n}(D)$ (4.17). This procedure leads to a finite increasing sequence of temperatures

$$\alpha_1 = 0 > \alpha_2 > \ldots > \alpha_K,$$

and a finite decreasing sequence of dimensions associated to the temperatures

$$D_1 = n > D_2 > \ldots > D_K = 1,$$

where for $i = 2, \ldots, K$

$$\alpha_i = \min_{j<D_{i-1}} \frac{\gamma_n(\hat{s}_j) - \gamma_n(\hat{s}_{D_{i-1}})}{f_n(D_{i-1}) - f_n(j)} = \frac{\gamma_n(\hat{s}_{D_i}) - \gamma_n(\hat{s}_{D_{i-1}})}{f_n(D_{i-1}) - f_n(D_i)}.$$

The passage to the minimal penalty is expected to be marked by a sudden fall of the dimensions of the associated partitions, phenomen observed in practice. The

method consists of selection the value $\hat{\alpha}$ of $(\alpha_i)_{1\leqslant i\leqslant K}$ associated to the highest jump of dimensions observed in $(D_i)_{1\leqslant i\leqslant K}$.

4.2 Calibration of the method

Here are presented some important practical problems on the estimation of α which can have consequence on the performance of the final estimator. A calibration of the method, which is expected to perform well in most situations, is proposed.

1. If the maximal jump of dimension of the sequence $(D_i)_{i=1,\dots,K}$ is attained by several values of $(\alpha_i)_{i=1,\dots,K}$, which one is to be chosen?
2. Does the maximal dimension of the considered partitions D_{max} affect the choice of α?

Among several values of α reaching the maximal jump, we decide to consider the smallest one since it is associated to the first abrupt change of dimensions. This choice appears to be reasonable from simulation not reported here.

In practice, some big dimension jumps are observed for too small values of α and a partition of too high-dimension can be selected. This is the situation in case (a) : in Figure 4 (a), the values a_1 and a_2 are the values of α respectively associated to the two biggest jumps. If $D_{max} = 40$, then $\hat{\alpha} = a_1$ and $\hat{D} = 15$, while if $D_{max} = 25$, then $\hat{\alpha} = a_2$ and $\hat{D} = 4$. To answer the second question, D_{max} can play a role in the selection of the penalized estimator. In practice, the user will have to choose D_{max} according to information he get. But our objective here is to propose an automatic method. A universal value of D_{max} is not relevant since it depends on the problem. We propose to eliminate the high-dimensionnal partitions by forcing α to be bigger than a noted threshold α_{thr}. Since in this framework $\alpha = \sigma^2$, we put $\alpha_i = \sigma^2\beta_i$. Then it suffices to choose β_{thr} and put $\alpha_{thr} = \sigma^2\beta_{thr}$. Since the variance σ^2 is unknown, it is substituate by one of its estimators. Since the number of change-points as well as their locations are unknown, the classical regression estimator cannot be used. One can find in the literature several good estimators of the variance ([15], ...). We decide here to take the estimator proposed in Hall *et al.* [11] which can be quickly computed:

$$\hat{\sigma}^2 = (n-M)^{-1}\sum_{k=1}^{n-M}\left(\sum_{j=0}^{M} d_j y_{j+k}\right)^2, \tag{4.19}$$

such that $\sum_{j=0}^{M} d_j = 0$ and $\sum_{j=0}^{M} d_j^2 = 1$. Following Hall *et al.*, we chose $M = 3$ and the associated sequence is

$$(d_j)_{j=0,\dots,3} = (0.1942, 0.2809, 0.3832, -0.8582).$$

Choice of the threshold β_{thr}. We propose to choose the threshhold β_{thr} from a null hypothesis test H_0 : "no change-point is present" against the alternate hypothesis H_1 : "there exists at least a change-point". Let β_v be the first value of the sequence $(\beta_i)_{i=1,\dots,K}$ for which the partition of dimension 1 is selected. The test is

$$H_0 : "\beta_{thr} \geq \beta_v" \quad \text{against} \quad H_1 : "\beta_{thr} < \beta_v".$$

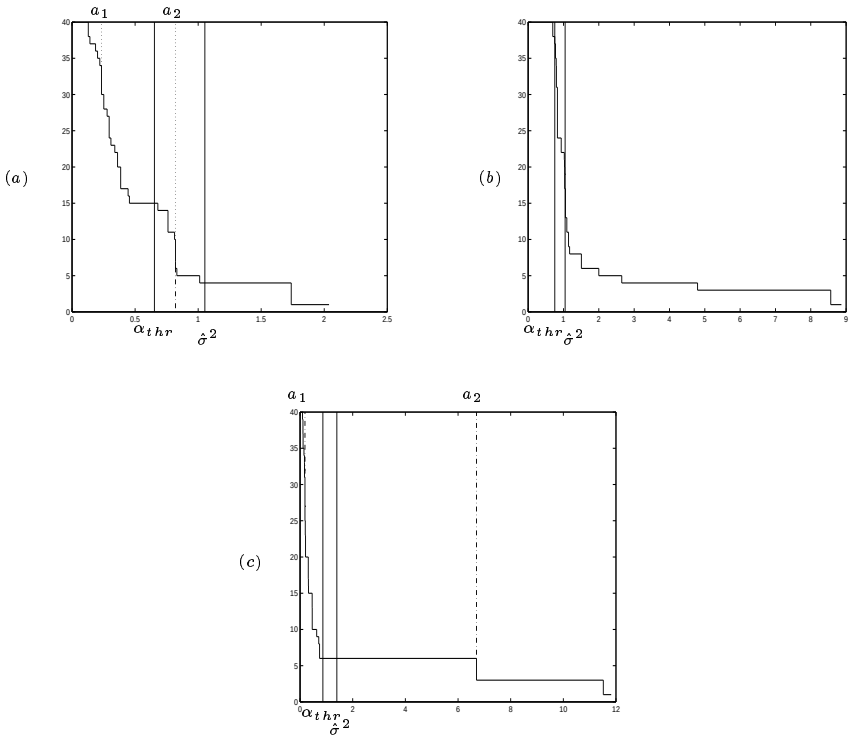

Figure 4: Function $\alpha \to \hat{D}_\alpha$ for the three realizations.

Let z be the level of the test,

$$\beta_{thr} = q_{1-z}(\beta_v),$$

where $q_{1-z}(\beta_v)$ is the quantile of order $1-z$ of β_v. It is estimated as its empirical version obtained on 2000 simulations. This procedure is performed 10 times for each n and the threshold $\beta_{thr}(n)$ is the average of the resulting 10 values. We consider two values of z according to $n < 200$ and $n \geq 200$ since the behavior of the function $\alpha \to \hat{D}_\alpha$ shows a certain stability for large n. The results are given in Table 2.

n	$\beta_{thr}(n)$	z
20	0.611	0.15
40	0.625	0.15
60	0.62	0.15
100	0.603	0.15
200	0.762	0.05
300	0.743	0.05
500	0.7	0.05
1000	0.714	0.05

Table 2: Estimations of $\beta_{thr}(n)$ for each n and two different test levels z.

It seems to be reasonnable to choose

$$\beta_{thr} = \begin{cases} 0.62 & \text{if } n < 200 \\ 0.76 & \text{if } n \geq 200. \end{cases}$$

An upper bound for α. In practice, it happens that α_{thr} is too large: σ^2 is overestimated, either β_{thr} is too large, and the minimal penalty cannot be detected. This is the case of the realization $y(c)$ (see Figure 4 (c)). By applying the method with threshold , $\hat{\alpha} = a_2$, the partition of dimension 1 is selected. To avoid this problem, α is bounded by $\hat{\sigma}^2$.

Finally, the calibrated method can be summarized as follows: the constant α is restricted to $[\alpha_{thr}, \hat{\sigma}^2]$ and

- in the case where there exists at least a jump of dimension,

$$\hat{\alpha} = \underset{\{i=1,\dots,K; D_{i+1}-D_i>0\}}{\operatorname{argmax}} \{D_{i+1} - D_i\} \tag{4.20}$$

- else

$$\hat{\alpha} = \alpha_{thr}. \tag{4.21}$$

Applications. Let us come back to the three realizations. The calibrated and the nocalibrated methods are respectively denoted CM and NCM. For comparison, the method with known and estimated variance (for which σ^2 in the penalty is substituted by $\hat{\sigma}^2$ (4.19)), noted respectively VC and NVC, are applied for the three realizations described in Subsection 4.1.2. In Figure 5 are plotted the estimators selected by each method and Table 3 gives the dimensions of these estimators with their associated loss function (2.3).

		CM	NCM	VC	NVC	$\inf_{D\geq 1} \|s - \hat{s}_D\|_n^2$
$y(a)$	$\hat{D}$	4	15	1	1	4
	$\|s - \hat{s}_{\hat{D}}\|_n^2$	0.196	0.643	0.386	0.386	0.196
$y(b)$	$\hat{D}$	6	6	5	5	6
	$\|s - \hat{s}_{\hat{D}}\|_n^2$	0.165	0.165	0.208	0.208	0.165
$y(c)$	$\hat{D}$	1	1	1	1	1
	$\|s - \hat{s}_{\hat{D}}\|_n^2$	0.142	0.142	0.142	0.142	0.142

Table 3: Dimensions selected by the calibrated and the nocalibrated methods, and the methods with known and estimated variance with their loss. The last column provides the best one in terms of loss.

For these particular realizations, the calibrated method works well since it selects the minimal loss estimator. For $y(a)$, since $a_1 < \alpha_{thr} < a_2$, CM selected $\hat{s}_4$, and for $y(c)$ the upper bound for α allows the method to select $\hat{s}_6$. Moreover, this calibrated method works better for $y(a)$ and $y(b)$ than the method with known and estimated variance.

5 Simulation experiments

In this section, some simulations experiments are performed to assess the performance of the calibrated proposed method. The first one compares the performance of some estimators selected by this method to the ones selected by other penalized criteria such that the C_p Mallows criterion and the BIC criterion. The second one proposes to assess the behavior of the method when the assumptions of the model are not valid. In the first case, s is not a piecewise constant function and in the second case, the noise is not Gaussian.

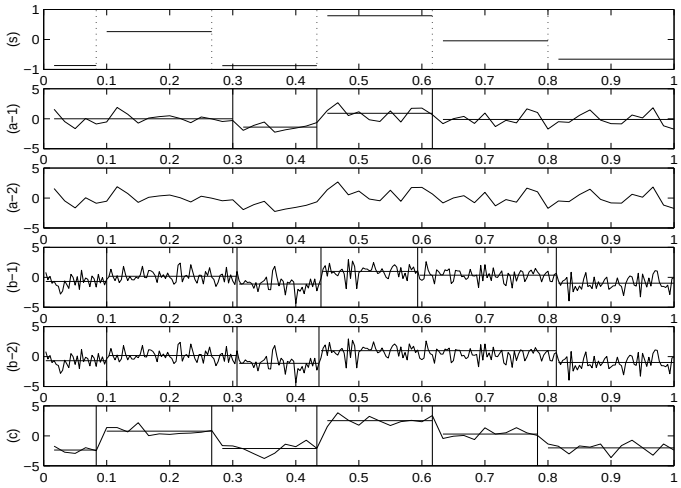

Figure 5: Penalized estimators selected by the calibrated method $(a-1)$, $(b-1)$ et (c) and the ones selected with known and estimated variance $(a-2)$, $(b-2)$ et (c).

5.1 Comparison with other criteria

The studied criteria are penalized least-squares criteria with the different following penalties :

- C_p Mallows
$$pen_{C_p}(D) = 2\sigma^2 \frac{D}{n}.$$
- BIC
$$pen_{BIC}(D) = \sigma^2 \frac{D}{n} \log n.$$
- The penalty obtained in our framework given by (2.9) with known variance
$$pen_{P_{\sigma^2}}(D) = \sigma^2 \frac{D}{n} \left(2 \log \left(\frac{n}{D} \right) + 5 \right).$$
- The penalty with estimated variance (given by (4.19))
$$pen_{P_{\hat{\sigma}^2}}(D) = \hat{\sigma}^2 \frac{D}{n} \left(2 \log \left(\frac{n}{D} \right) + 5 \right).$$
- The calibrated penalty
$$pen_{P_{CM}}(D) = \hat{\alpha} \frac{D}{n} \left(2 \log \left(\frac{n}{D} \right) + 5 \right).$$
where $\hat{\alpha}$ is obtained by the calibrated method.

We consider $\sigma^2 = 1$, $n = 60, 300$ and six different functions s (numbered from s_1 to s_6) plotted in Figure 6. Proceeding as in Section 3, for each n and each function s, the risk ratio $F_n(s, 2, 5)$ (3.14) is estimated over 500 simulations and the percentage of the number of times, denoted $\%p_{min}$, that the considered criterion leading to the

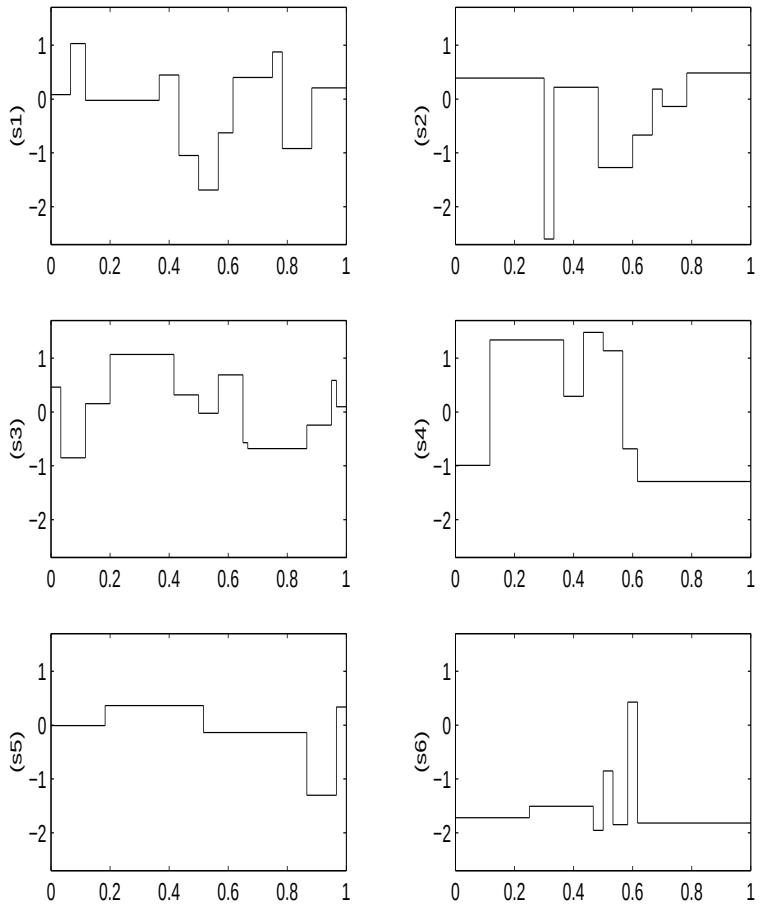

Figure 6: Fonctions s_i $i = 1, \ldots, 6$.

minimal loss estimator over the 500 simulations is calculated. The results are given in Table 4. Moreover, in Figures 7 and 8 are represented the distribution of $\hat{D}$ for the different s, n and for each penalty function. The two last column correspond respectively to the dimension realizing $O_{(n,r)}(s, \mathcal{S})$ (3.13), denoted $\hat{D}_O$, and the true one, denoted D_T.

Some comments are in order:

- The C_p Mallows works really bad, particulary for large n. It tends to under-penalize and selects partitions with too large dimensions. The $\log n$ term in the other penalties improves dramatically the results. The C_p Mallows is not suitable for selecting a partition from a large collection of partitions (as it is suspected in theory [3]).
- The risk ratios associated with P_{σ^2}, $P_{\hat{\sigma}^2}$ and P_{CM} are smaller than the one associated with BIC whatever n and s. The BIC criterion selects partitions with larger dimensions. The penalty function defined by (2.9) with $c_1 = 2$ and $c_2 = 5$ appears to be more accurate than BIC.
- For the penalty P_{σ^2}, $P_{\hat{\sigma}^2}$ and P_{CM} the risk ratios are close and tends to 1 with n : the estimators approach the best one when n increases.

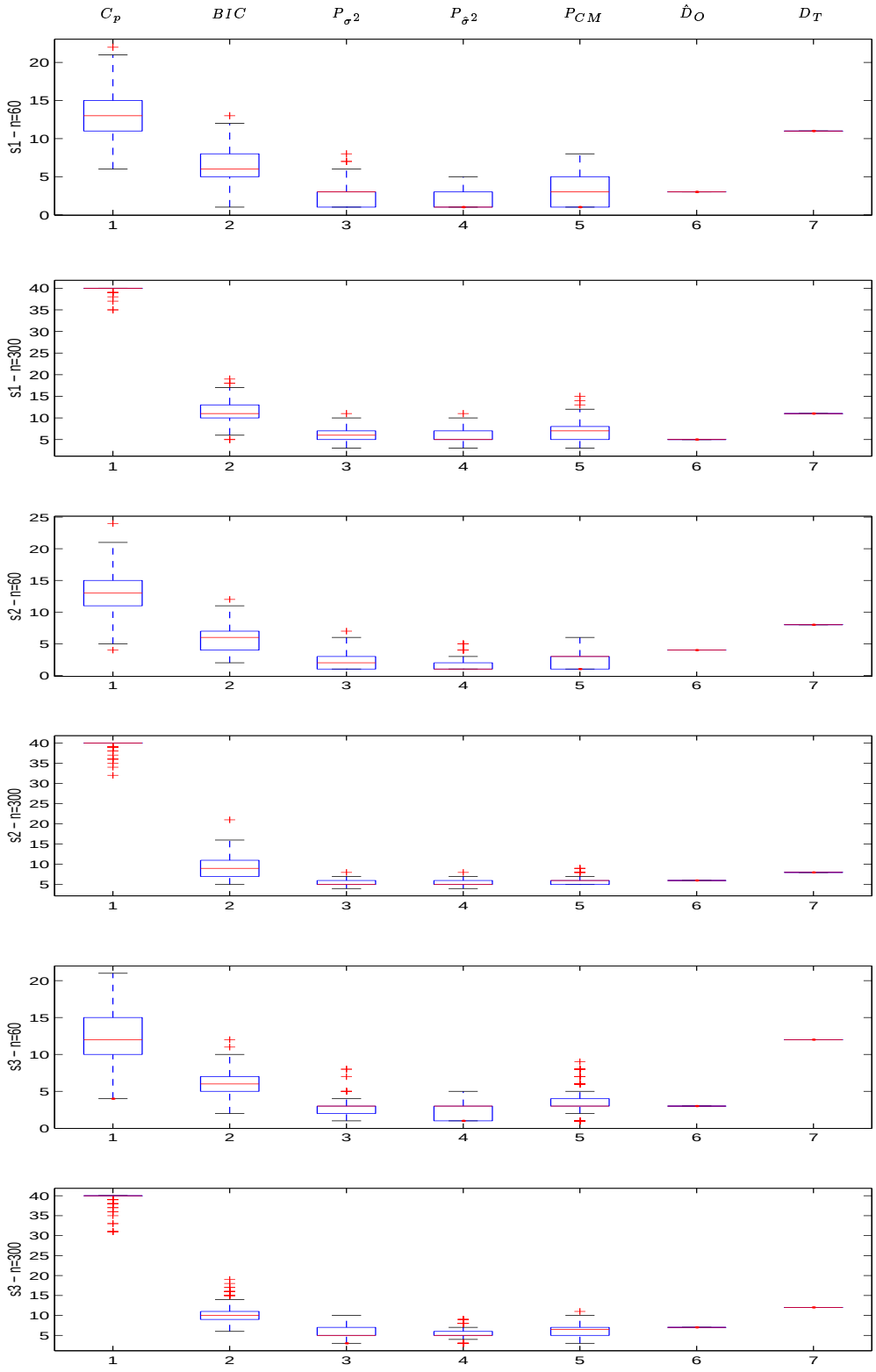

Figure 7: Distribution of $\hat{D}$ for s_1, s_2 and s_3, for each penalty function. The two last column correspond respectively to the dimension realizing $O_{(n,r)}(s, \mathcal{S})$ (3.13) and the true one.

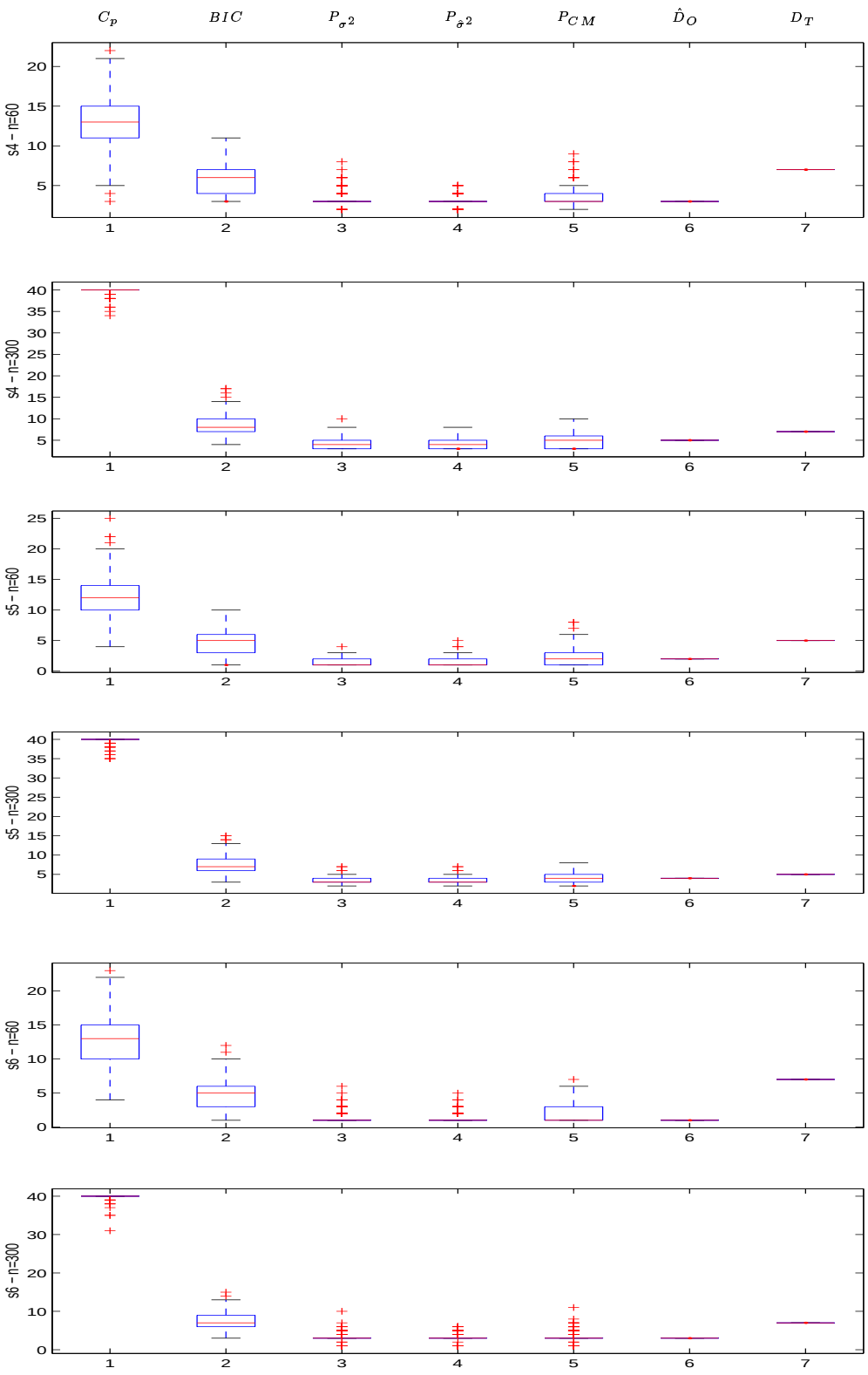

Figure 8: Distribution of $\hat{D}$ for s_4, s_5 and s_6, for each penalty function. The two last column correspond respectively to the dimension realizing $O_{(n,r)}(s, \mathcal{S})$ (3.13) and the true one.

		$n = 60$					$n = 300$				
		C_p	BIC	P_{σ^2}	$P_{\hat{\sigma}^2}$	P_{CM}	C_p	BIC	P_{σ^2}	$P_{\hat{\sigma}^2}$	P_{CM}
s_1	$F_n(s_1,2,5)$	1.73	1.22	1.15	1.17	1.1	3.05	1.24	1.003	1.005	1.003
	$\%p_{min}$	0.2	12.8	18.2	16.6	21.6	0	5.2	22	21.8	21.2
s_2	$F_n(s_1,2,5)$	2.18	1.32	1.18	1.3	1.16	5.8	1.76	1.06	1.07	1.08
	$\%p_{min}$	0.2	13.8	20.8	11.4	20.2	0	10	52.2	50.4	52.4
s_3	$F_n(s_1,2,5)$	1.9	1.27	1.13	1.15	1.1	3.3	1.26	1.08	1.09	1.03
	$\%p_{min}$	0.4	12	22.6	19.6	23.6	0	9.8	17.8	18.4	26.2
s_4	$F_n(s_1,2,5)$	2.9	1.8	1.13	1.07	1.16	5	1.52	1.06	1.06	1.05
	$\%p_{min}$	0.4	13.8	69	74.6	62.8	0	7.4	27	26.6	31.4
s_5	$F_n(s_1,2,5)$	3.1	1.74	1.19	1.19	1.22	9	1.9	1.16	1.16	1.13
	$\%p_{min}$	0	13	28.8	23.8	26.2	0	6.2	30	30	33.2
s_6	$F_n(s_1,2,5)$	3.19	1.78	1.07	1.05	1.092	8.16	2.26	1.1	1.13	1.16
	$\%p_{min}$	0	12	40.8	39	39.4	0	6.6	69.8	69.6	60.8

Table 4: For each function s_i and each n, estimation of the risk ratio of the penalized estimator obtained by each criterion and percentage of the number of times that the considered criterion leading to the minimal loss estimator over the 500 simulations.

- The calibrated method can do better than the method with known and estimated variance, even if the risk ratios are very close. The calibrated method seems to correct the constant penalty in order to select the minimal estimator in terms of risk. The constants $c_1 = 2$ and $c_2 = 5$ have been chosen to be optimal in most situations and can be suboptimal for specific n and s values.
- Figures 7 and 8 show that on these particular exemples the minimal risk dimension, $\hat{D}_O$, is not equal to the true one D_T. The BIC criterion seems to select a dimension close to D_T while P_{CM} a dimension close to $\hat{D}_O$ which is the aim here. That confirms the difference between the aim here which is to select the minimal risk estimator and a asymptotic approach in which the estimator tends to the true.

5.2 Case of a function s no piecewise constant

Here we want to answer to the following question : what happens if the function s does not belong to no $\mathcal{S}_m$?

Two realizations are simulated with $n = 1000$, a variance $\sigma^2 = 0.1$ and from the functions denoted by s_7 and s_8 and plotted respectively in Figures 9 (a) (left) and 9 (a) (right). The two penalized estimators selected by the calibrated method are respectively plotted in Figures 9 (b) (left) and 9 (b) (right) with the realizations. Note that these estimators realize the minimal loss.

These results suggest that the method leads to estimators giving good approximation performances.

5.3 Case of a no Gaussian noise

Now the method is applied when the signal noise is not Gaussian. Two different noises were considered:

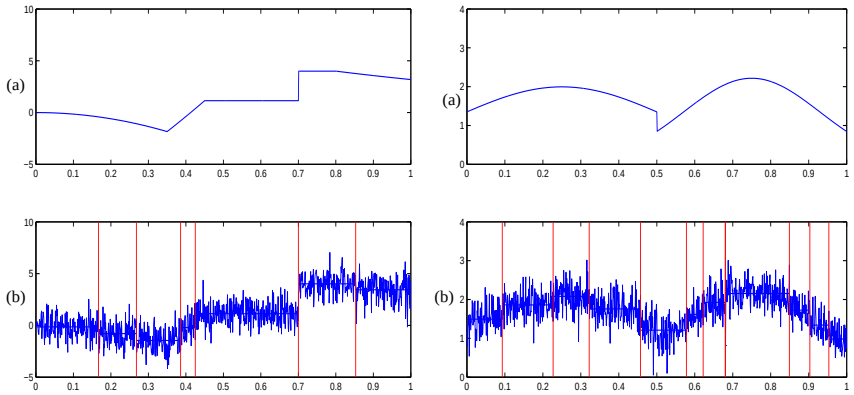

Figure 9: Fonction s_7 (a) and penalized estimator (b) (left) - fonction s_8 (a) and penalized estimator (b) (right).

- ε is a Laplace (symetric Exponential) noise (marked by Lap).
- ε is a Bernoulli noise with parameter $1/2$ (marked by Ber).

Recall that under the Gaussian assumption, the penalty form comes from a control of a Gaussian process. By comparing the Laplace distribution to the Gaussian distribution in the tail, one can suspect that for Laplace, a stronger penalty should be prefered to get a similary control. For the Laplace noise, the estimator associated to the double of the penalty estimated by the calibrated method is considered (marked by $Lap-2$) since from simulations it seems to be more appropriate.

We consider $n = 500$, and five simulated constant functions s. For each noise, the risk ratio (3.14) is estimated over 500 simulations. The results are given in Table 5.

	s_1	s_2	s_3	s_4	s_5
Lap	2.56	2.15	2.63	2.4	1.89
$Lap-2$	1.29	1.34	1.5	1.13	1.1
Ber	1.0008	1.004	1.002	1.0005	1.0003

Table 5: Risk ratio for different functions s and different noise.

For the Bernoulli noise, the risk ratio is close to 1 whatever s, the method works well since the noise is small and the change-points are marked. It is not the case for the Laplace noise (Lap) and a correction of its penalty decreases the risk ratio ($Lap-2$) as suspected. The penalty is too small to penalize correctly (the constants are certainly not adapted), one underpenalizes and a partition of too large dimension is selected.

6 Application : detection of the changes in the monthly number of tests HIV in France

Data considered here are the number of tests HIV executed every month in France between February 1987 and October 1991 (see Figure 11). The data are supposed

to be Gaussian and independent. We want to detect change-points in the mean revealing some changes in the behavior of French people facing the virus.

The function $\alpha \to \hat{D}(\alpha)$ is represented in Figure 10. By applying the calibrated method, the dimension 13 is selected but the biggest jump, which is associated to a_2, does not belong to $[\alpha_{thr}, \hat{\sigma}^2]$ and the minimal penalty is not considered. This jump leads to the estimator of dimension 3 plotted in Figure 11. That reveals three time intervals of constant mean of the number of tests with a clear increase between October 1991 and August 1995. This period seems to correspond to the conscience hold of the seriousness of the virus. Next a light decrease is observed.

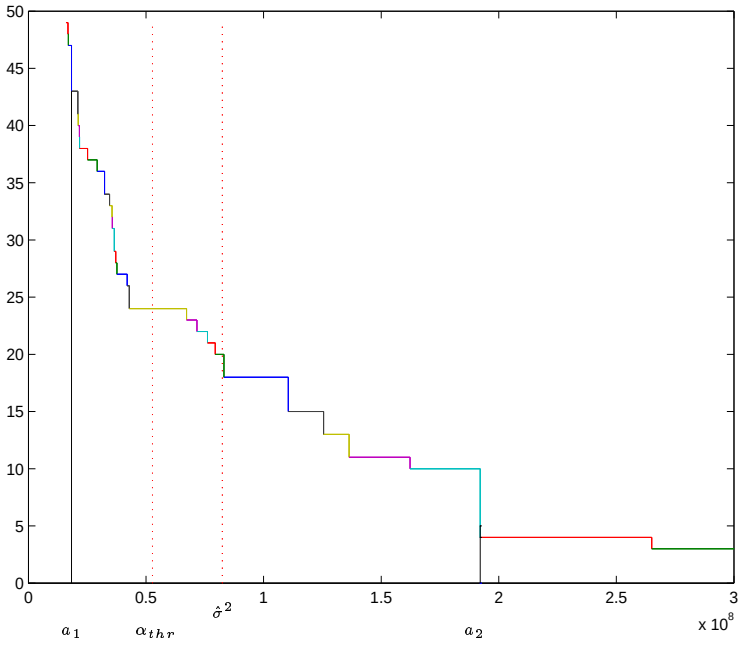

Figure 10: Function $\alpha \to \hat{D}(\alpha)$.

Moreover a second big jump of dimension, associated to a_1, is present (see Figure 10). This jump will be never considered by the method since it is smaller that the one associated to a_2. However, it could be interesting to look for its associated estimator which has the dimension 27 and is plotted in Figure 11. That brings a more precise segmentation and so an additionnal information. It seems to reveal the existence of an annual cycle showing an increase between the beginning of spring and the end of summer.

7 Discussion

We have proposed a calibrated penalized least-squares criterion for the change-points problem via a nonasymptotic approach. This leads to an automatic method which works well and better that the C_p Mallows and the BIC criteria particulary for small samples. However it is difficult to obtain a calibration working well in all situations and in practice it could be interesting to take into account the information available from an user. Indeed, as we have seen in the previous application,

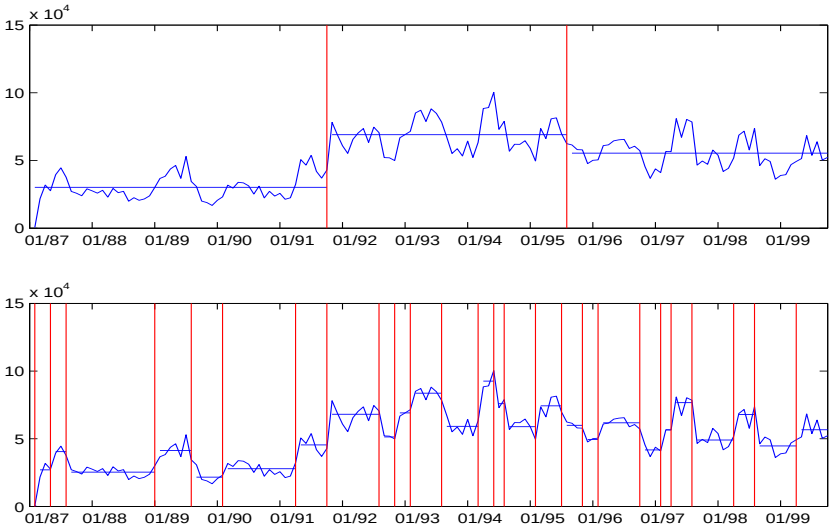

Figure 11: Penalized estimators of dimensions 3 and 27.

the calibration can be prove to be too stringent in the sense that the true minimal constant penalty does not belong to the interval of calibration $[\alpha_{thr}, \hat{\sigma}^2]$. For one application, the user can perform the nocalibrated method and consider the estimator associated to the biggest jump, and may be the different ones associated to different big dimension jumps if any. A difficulty occurs when the biggest jump is too small to be considered as signifiant. This situation can be explained by the fact that the two constants $c_1 = 2$ and $c_2 = 5$ are not the optimal ones for this particular application. Moreover, in practice one can remark that a multiplicative constant different but close of 2 can work better. Indeed $2\hat{\alpha}$ can be close to one of values in $(\alpha_i)_{i=1,\dots,K}$ and could be interesting in this kind of configuration to be considered.

One of the aim of this study was to assess the heuristic method proposed by Birgé and Massart for the penalty constant estimation, and to calibrate it in the particular Gaussian regression framework. Numerical experiments have clearly shown that the calibrated proposed method performs well in terms of quadratic risk. That allows to hope it will be useful in more complex situations in the sense that the penalty constant α is not explicitly known, as for example in the CART framework [10].

References

[1] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pages 267–281. Akadémiai Kiadó, Budapest, 1973.

[2] M. Basseville and N. Nikiforov. *The Detection of abrupt changes - Theory and applications*. Prentice-Hall: Information and System sciences series, 1993.

[3] L. Birgé and P. Massart. Gaussian model selection. *J. Eur. Math. Soc.*, 3:203–268, 2001.

[4] L. Birgé and P. Massart. A generalized C_p criterion for Gaussian model selection. Technical report, Publication Université Paris-VI, 2001.

[5] L. Birgé and Y. Rozenholc. How many bins should be put in a regular histogram. Technical report, Publication Université Paris-VI, 1999.

[6] B.E. Brodsky and B.S. Darkhovsky. *Nonparametric methods in change-point problems.* Kluwer Academic Publishers, the Netherlands, 1993.

[7] G. Castellan. Sélection d'histogrammes à l'aide d'un critère de type akaike. *C. R. Acad. Sci., Paris, Sér. I, Math. 330*, 8:729–732, 2000.

[8] A. Chambaz. Detecting abrupt changes in random fields. *ESAIM*, 6:289–309, 2002.

[9] H.G. Müller E. Carlstein and D. Siegmund. *Change-point problems.* Hayward, CA : Institute of Mathematical Statistics. Papers from the AMS-IMS-SIAM Summer Research Conference held at Mt. Holyoke College, South Hadley, MA, July 11-16, 1992.

[10] S. Gey and E. Nedelec. Model selection for CART Regression Trees. Technical report, Publication Université Paris-XI, 2001 - (A revised version is to appear in IEEE Transaction and Information Theory).

[11] P. Hall, J.W. Kay, and M.D. Titterington. Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, 77:521–8, 1990.

[12] M. Hanke. Limitations of the L-curve method in ill-posed problems. *BIT*, 36(2):287–301, 1996.

[13] P.C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, 34(4):561–580, 1992.

[14] S. M. Kay. *Fundamentals of statistical signal processing - Detection theory*, volume II. Prentice Hall signal processing series, 1998.

[15] M. Lavielle and E. Lebarbier. An application of MCMC methods for the multiple change-points problem. *Signal processing*, 81:39–53, 2001.

[16] M. Lavielle and C. Ludena. The multiple change-points problem for the spectral distribution. *Bernoulli*, 6(5):845–869, 2000.

[17] M. Lavielle and E. Moulines. Least Squares estimation of an unknown number of shifts in a time series. *Jour. of Time Series Anal.*, 21:33–59, 2000.

[18] E. Lebarbier. *Quelques approches pour la détection de ruptures à horizon fini.* PhD thesis, Université Paris XI, 2002.

[19] F. Letué. *Modèle de Cox : estimation par sélection de modèle et modèle de chocs bivarié.* PhD thesis, Université de Paris XI, 2000.

[20] C.L. Mallows. Some comments on Cp. *Technometrics*, 15:661–675, 1974.

[21] B. Q. Miao and L. C. Zhao. On detection of change points when the number is unknown. *Chinese J. Appl. Probab. Statist.*, 9(2):138–145, 1993.

[22] G. Schwarz. Estimating the dimension of a model. *Ann. Stat.*, 6:461–464, 1978.

[23] Y.C. Yao. Estimating the number of change-points via Schwarz criterion. *Stat. & Probab. Lett.*, 6:181–189, 1988.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Futurs : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399